Cold Spring Harbor Laboratory Meeting "Telomeres & Telomerases" 2-6 May 2007

TELOMERES IN EVOLUTION AND DEVELOPMENT FROM BIOSEMIOTIC PERSPECTIVE

Guenther Witzany
Telos – Philosophische Praxis
Vogelsangstrasse 18c, A-5111-Buermoos, Austria
Tel./Fax: ++43/6274 6805
Mail: witzany@sbg.at

Abstract:

Whereas Telomeres protect terminal ends of linear chromosomes telomerases identify natural chromosome ends being different from broken DNA. Although telomeres play a crucial role in the linear chromosome organisation of eukaryotic cells, their molecular syntax descended from an ancient retroviral competence. This is an indicator for the early retroviral colonization of large double stranded DNA viruses, which are putative ancestors of the eukaryotic nucleus.

This contribution will demonstrate an advantage of the biosemiotic approach towards our evolutionary understanding of telomeres: focus on the genetic/genomic structures as language-like text which follows combinatorial (syntactic), context-sensitive (pragmatic) and content-specific (semantic) semiotic rules. Genetic/genomic organisation from the biosemiotic perspective is not seen any longer as an object of randomly derived alterations (mutations) but as functional innovation coherent with the broad variety of natural genome editing competences of viruses.

Keywords: telomeres, telomerases, eukaryotic nucleus, viruses

1. Introduction

Biosemiotics investigates both communication processes within and among cells, tissues, organs, organisms as sign-mediated interactions and nucleotide sequence order as codes/texts which follow syntactic, pragmatic and semantic rules. In the latter case biosemiotics investigates genetic sequences as codes/texts which are coherent with laws of physics and chemics, but additionaly follow a complementary mix of combinatorial (syntactic), context-sensitive (pragmatic), content-specific (semantic) rules. In this respect it is interesting from biosemiotic perspective to ask some questions about the roles of telomeres and telomerases in evolution, structure and content arrangement of genomes: What does the telomere specific "molecular syntax" (Eigen and Winkler 1975) indicate? Why do we find telomeres as key features mainly in the context of linear chromosomes of eukaryotes not in circular prokaryotic genomes? What's the evolutionary relation of non-mobile telomeres and mobile genetic elements like telomerase and other reverse transcriptases? How can we understand also a telomerase-similar activity being capable for telomere elongation in several insects and plants?

Because telomeres and telomerases are key features of eukaryotes *and* some viruses (which may inhabit prokaryotes) it will be necessary to describe the crucial differences between eukaryotes and prokaryotes which will show that the eukaryotic nucleus doesn't

derive from prokaryotes but from a assembly of viral competences. To understand the evolutionary context it is necessary to describe viral key features and their possible role in the evolution of life.

From biosemiotic perspective it will be interesting to clarify the relations between molecular syntax of telomere repeats with its meaning, i.e. function in the genomic content. Therefore it will be necessary to have a look at its evolutionary roots. In this context the telomere replication process by telomerase is the most interesting feature because it is processed by a very ancient competence (Nosek et al. 2006), i.e. reverse transcriptase with a great variety of functions throughout key processes of living nature (Eickbush 1997). A coherent description of this typical characteristics of telomeres and telomerases in a biosemiotic view include a holistic view on their molecular syntax, their in vivo - meaning function (semantics) and their pragmatic relevance for concrete genome editing agents.

2. Function of telomere repeats

If we look at the specific characteristics of telomeres we find some features which are common to all genomes which possess telomeres: Telomeres are highly conserved non-mobile repetitive DNA-sequences. Telomeres are nucleoprotein structures which protect the ends of chromosomes from erosion, degradation, colonization, or sticking together chromosome ends (Blasco 2007). They are necessary only in linear chromosomes not in circular. Telomere repeats are building nodes. These nodes stabilize telomeres and *are not of linear DNA*. These nodes also care for not being recognized as DNA damage which would induce a DNA repair pathway. Is the node intact this serves as signal for the cell that she is fit for further replications. Telomeres are suggested to be the forerunners of centromeres which derived most likely from an ancestral telomere-telomere fusion (Ijdo et al. 1991). Similar to telomeres centromeres are highly conserved non-mobile repetitive DNA-sequences. They interact with spindle microtubules and are therefore crucial for distribution of chromosomes to offspring cells (Villasante et al. 2007) as well as they encode small RNAs which care for heterochromatin formation (Couzin 2002, Grewal and Elgin 2007).

Linear chromosomes of eukaryotes have the so-called end-replication problem: DNA Polymerases which replicate leading strands of doublestranded DNA only in 5’ to 3’ direction are not able to replicate lacking strands, i.e. 3’ to 5’ direction. For leading strand replication DNA Polymerases are adding polynucleotides to a RNA primer. These RNA strands are later replaced by DNA. At the terminal end of the chromosome the RNA primer cannot replaced completely by DNA, so it cannot code for proteins or further replications. When the last RNA is added, DNA polymerase and DNA ligase come to transform the RNA of the primer to DNA. But for this process there must be another DNA strand in front of the RNA primer. The end-replication problem is that there is no other DNA strand in front of the last attached RNA primer. That RNA is degraded by enzymes. Thus, a section of telomeres would be lost during each cycle of replication. For complete lacking strand replication another technique is necessary. A reverse transcriptase named telomerase uses its integrated subunit, an inherent RNA template for replication of the overhanging RNA primer so that the terminal end of the lacking strand can be fully completed without loss (Haoudi and Mason 2000).

Telomere function needs a certain length. If this length is undergone by continued end-replication problems or damage the protection of chromsome ends doesn’t function (Du and Traktman 1996). Therefore it has been suggested that continous telomere shortening is a main reason for cell aging. Recent research documents that this is the case only in vitro experiments not in vivo. (Laun et al. 2007).

Nature Precedings : doi:10.1038/npre.2007.932.2 : Posted 11 Nov 2007

2.1. Differences in the molecular syntax of telomere sequences

From biosemiotic perspective it would be interesting whether the telomere sequences differ among various organisms, species, kingdoms, because this could be an indicator that primarily the function of telomere repeats is of importance, not the sequence order that encodes these function.

Interestingly the molecular syntax of telomere repeats differs in various organisms where it has been identified. This indicates that there is no unique molecular syntax necessary to guarantee the function which telomere repeats have to fulfil, but in difference to this, that the same important function can be coded by different nucleic acid sequences. For example we find TTAGGG in vertebrates, humans, mouse, xenopus, filamentous fungi, neurospora crassa, slime moulds physarum, didymium; TTGGGG in tetrahymena, glaucoma; TTGGG(T/G) in paramecium; TTTTGGGG in oxytricha, stylonychia, euplotes; TTAGGG(T/C) in apicomplexan protozoa plasmodium; TTTAGGG in *Arabidopsis thaliana*; TTTTAGGG in green algae chlamydomonas; TTAGG in insects *Bombyx mori*; TTAGGC in roundworms ascaris lumbricoides; TTAC(A)(C)G(1-8) in fission yeasts schizosaccharomyces pombe; TGTGGGTGTGGTG (from RNA template) in saccharomyces cerevisiae; GGGGTCTGGGTGCTG in candida glabrata; or GGTGTACGGATGTCTAACTTCTT in candida albicans.

Telomeres act as immune functions against genomic agents with high recombination- or degradation- competences, i.e. viral genetic parasites, and function similar to an RNAi system. RNAi protects the genome against genomic parasites, i.e. viruses, by silencing genomic transcripts of exogenous infective RNA viruses or endogenous transposons or retroposons (Fire et al. 1998, Couzin 2002). Additionally telomeres serve as recognition sequences, primer functions and genetic/genomic raw material for sequence generation (genome duplication, RNA template).

In Drosophila and some plants telomere elongation during replication does not occur by telomerase but through recombination facilitated by non-LTR retroposons HetA and TART (Nakamura and Cech 1998, Blasco 2007, Fajkus et al. 2005). They transport their *gag* protein into nucleus to produce more copies to the chromosome ends (Rashkova et al. 2002). These retroposons which fullfil the same function of telomere elongation as telomerase are regulated by the same epigenetic regulations that govern mobile elements activity including also RNAi (Savitsky et al. 2006, Slotkin and Martienssen 2007).

3. Telomere replication in most cases by telomerase a reverse transcriptase

In most cases except that described before in certain arthropods and plants telomeres are replicated by telomerase, a reverse transcriptase. This indicates that the function of telomerases in eukaryotic replication cycle is very ancient (Curcio and Belfort 2007). Some authors suggest that reverse transcriptases derived from RNA dependend RNA Polymerases which descended from an ancient RNA world (Boeke 2003).

Telomerase is a ribonucleoprotein enzyme that is an assembly of telomerase RNA and telomerase reverse transcriptase (Jady et al. 2004). Telomerase is clearly related to mobile elements especially to the non-LTR-retroposons (Eickbush 1999).

3.1. Reverse transcriptases and mobile elements

Mobile Elements in the genome may be transposons which integrate directly into a host genome or retroposons which integrate via reverse transcriptase. Copying from RNA into DNA generally occurs through reverse transcriptase. Mobile elements are important for genotype processing with far reaching consequences for phenotype expression during its

various developmental stages. Recent research demonstrates that overlapping epigenetic marking in eukaryotic cells is an important evolutionary feature to silence the expression of mobility of these mobile elements (Slotkin and Martienssen 2007). Mobile elements are able to silence single genes as well as larger chromosomal regions and play therefore an important role in the evolution of diversity. Their competence to recombine, rearrange and insert into genomic content they share with retroelements (Coffin et al. 1997). They influence neighbouring genes through alternative splicing and are active agents as enhancers, promotors or act by polyadenylation patterns (Slotkin and Martienssen 2007).

Reverse transcriptases play key roles in mobile elements like transposons and retroposons. One type of retroposons has direct repeats on their ends (LTR) others not (non-LTRs). Interestingly their number increases with every transposition (transposition duplication) so that they can expand genomes: LINE-1 is 20% of the human genome (Maita et al. 2004). In opposition the transposon contains a code for the transposase protein. This enzyme identifies the terminal inverted repeats which flank mobile elements, excises it and integrates itself instead of the excised one. The gap at the donor site is repaired in a cut and paste transposition or filled up with a copy of the transposon by a gap repair technique (Slotkin and Martienssen 2007). Transposons can integrate themselves also in phages and plasmids and can be transferred with them in other cells (Frost et al. 2005).

In contrast to non-mobile telomeres and centromeres mobile sequences such as transposons and retroposons (Volff 2006) and non-protein-coding repetitive elements such as LTRs, SINEs and LINEs enable far-reaching DNA rearrangement and reorganization (Shapiro 2002, Sternberg 2002, Shapiro and Sternberg 2005). Together, they play a decisive role in the evolution of new genomic structures (Shabalina and Spiridonov 2004, Shapiro and Sternberg 2005, Sternberg and Shapiro 2005). Interestingly the non-protein-coding DNA contains also the regulations of transcription, promoter, enhancer and suppressor (Bird et al. 2006). The repetitive sequences are highly species-specific and more suitable for the determination of species than the coding sequences (Villarreal 2005).

3.2. Reverse transcriptases play major roles in natural genome editing

Additionally reverse transcriptases play key roles in altering genomic structures and therefore play an important role in evolutionary processes processed by natural genome editing (Witzany 2006). Reverse transcriptases generate (a) copies of mRNAs which they need for integration into a genome (b) copies of non-mRNAs like small nucleolar RNAs, one of the largest classes of non-coding RNAs (Zemann et al. 2006) which are as DNA copies SINEs. SINEs can initiate new genes which code for small RNAs with regulatory competences on existing genes. One further key feature of reverse transcriptases is that reverse transcriptase is a primer for retroposons such as LTRs (copia, gypsy, Ty1, IAPs, HERVs), non-LTRs act like telomerases in several arthropods: Het-A/TART, SINEs, LINEs, ORF1 (RNA-binding protein), ORF2 (endonuclease, reverse transcriptase activities), ALUs (manipulation of LINE 1 function for mobilization), Group II self splicing introns, snoRNAs (Type 1-3 Retroposons) with all their important regulatory functions (Yang et al. 1999, Batzer and Deininger 2002, Tomlinson et al. 2006, Weber 2006, Matera et al. 2007).

Reverse transcriptases are found also in retroviruses of mammals, birds, in Hepadnavirus of mammals and birds and Caulimovirus of plants, in LTR-Retroposons of animals, plants, fungi and protazoa, in non-LTR retroposons of animals, plants fungi and in protazoa, group II introns of bacteria, fungi, plant mitochondria, chloroplasts and plastids, in mitochondrial plasmids of Neurospora mitochondria and in multiple singlestranded DNAs (Villarreal 2005).

Reverse transcriptases together with RNA dependent RNA polymerases replicate positive strand RNA viruses, doublestranded RNA viruses, negative strand RNA viruses and

Retroviruses (Koonin et al. 2006). RNA dependent RNA polymerases are involved in the coupling of heterochromatin to the production of siRNAs (Sugiyama et al. 2005). The RNAi system is competent in posttranscriptional gene silencing and therefore a crucial instrument in keeping the balance between the need for expression and the need for silencing (Grewal and Elgin 2007)

As mentioned also open reading frames (ORFs) code for reverse transcriptase. RNA dependent DNA polymerase has relations to RNA dependent RNA polymerase. Many organisms have ORFs which code for proteins that have very similar sequences as retroviral reverse transcriptases (Xiong and Eickbush 1990, Mesnard and Lebeurier 1991). If we root these lines of descent in RNA dependend RNA polymerases we find 2 groups, (i) group 1 contains: LTR retroposons, RNA Viruses, DNA Viruses and (ii) group 2 contains: non-LTR-retroposons, bacterial and other organelle parts (Nakamura and Cech 1998).

But Telomerase function has alternatives: Not only Telomerase Reverse Transcriptase (TERT) replicate telomere-repeats but protein priming, terminal hairpins and recombination which allow complete replication of (i) viral linear DNA, (ii) bacterial plasmid genomes and (iii) linear mitochondrial genomes of certain eukaryotes (Nakamura and Cech 1998).

The telomerase function is cell-cycle regulated. Telomerase functions exclusively if its suppression is deleted. Is the telomerase function in telomere-replication fulfilled a signal initiates its suppression again. If this signalling process is disturbed uncontrolled cell replication may occur (with even fatal consequences). Telomerase has to be transported to telomere repeats for its elongation during S-phase of cell cycle. The delivery agents are Cajal bodies, small nucleolus-like organelles competent in (i) splicing, (ii) ribosome production and (iii) transcription (Platani et al. 2002, Jady et al. 2004) residing the periphery of nucleoli (Darzacq et al. 2004, Matera 2006). Cajal bodies undertake movements including the whole area of the nucleus, and for certain properties they fuse with other Cajal Bodies as well as they associate with nucleoli (Tomlinson 2005, Kiss et al. 2002). Telomerase trafficking is restricted to s-phase of cell cycle and to avoid telomerase activity at non-telomeric sites of the chromosomes. (Tomlinson et al 2005).

4. Eukaryotic key features not present in prokaryotes

Because telomere repeats are key features of eukaryotes only in rare cases of prokaryotes we may conclude, that eukaryotic telomeres and telomerases are interconnected with the evolution of the eukaryotic cells. If we want to decipher the evolutionary roots of telomeres and telomerases we have to look on the main differences of eukaryotes and prokaryotes. If we detect the evolutionary descent of the eukaryotic nucleus we may even find the roots of telomeres and telomerases. Therefore we should identify eukaryotic key features not present in most prokaryotes.

First, eukaryotic genomes share a great variety of repeat elements with higher order regulatory functions. Contrary to prokaryotes eukaryotic replication proteins have very different amino acid sequence compositions. Additionally eukaryotes share the control of DNA packaging and replication whereas prokaryotes doesn't have chromatin proteins like histones.

The eukaryotic DNA replication starts in numerous (thousands) of sites and is regulated by a complex cell cycle regulatory system. Eukaryotic replication control proteins doesn't have similarity to prokaryotic ones. A further difference between eukaryotes and prokaryotes is that daughter cells segregate by attachment to a microtubule system (spindles) not by attachment at the membrane. The highly conserved mitotic spindle system is not found in any prokaryote (Cottingham and Hoyt 1997).

Also the eukaryotic nucleus posses three classes of DNA dependent RNA polymerases that lacks similarity to polymerases of prokaryotes. A crucial difference is that in eukaryotes

the products of RNA polymerases must undergo posttranscriptional modifications (splicing) before they can function in the cytoplasm as mRNA, tRNA, rRNA. In no prokaryote splicing of pre-mRNAs is found. To prevent mistranslation of mRNA or unspliced tRNA the nucleus has to separate transcription/processing of mRNA from the cytoplasm transport of processed RNAs. Therefore a nuclear membrane is needed to segregate transcription, mRNA processing, transport and translation in the cytoplasm (Vale 2003). The nuclear membrane is distinct from plasma membrane and is dissolved after S-phase and reformed at late anaphase/telophase.

Introns allow splicing: Group I introns are mobile elements which code for DNA transposase, group II introns code for reverse transcriptase and small RNAs, recognize the splice junctions and splice RNA after capping, but all 3 intron types are not existent in prokaryotes (but in prokaryote viruses)

All complex modifications of mRNA und nuclear RNA seem to be acquired during Evolution of the eukaryotic nucleus and are highly conserved in eukaryotes but absent from prokaryotes.

Only very few prokaryotic genomes share some of these features. In the case of the spirochetes *Borrelia* they possess 3 types of telomeres, segmented genomes of linear and circular plasmids and extensive DNA rearrangements (Chaconas 2005, Tourand et al. 2006). As we will see this could be an indicator of intensive infection by competing genetic parasites which are in balance as an addiction module in a persistent status not harming the host but harmfull to organisms which lack this persistent status.

5. A viral progenitor of the eukaryotic nucleus?

The eukaryotic cell doesn't evolve by chance mutations in prokaryotic genomes over a long period of time but by a symbiogenetic integration of former free living bacteria. But this symbiogenetic integration can't explain the progenitor of the eukaryotic nucleus, because its key features could not derive from prokaryotes. He resembles a lot of key features, proteins and RNAs described above which are not found in prokaryotes. Interestingly they can be found in some prokaryote viruses (Villarreal 2005, Forterre 2006). These viruses use linear chromosomes, telomere repeats, multiple membranes, histone packaged chromosomes with marking effect for self/non-self identification and nuclear pores.

No single virus resembles all of these key features, but every key feature of the eukaryotic nucleus is present in large dsDNA-Virus. So we have to think about a process where different viral competences have been integrated into one single dsDNA virus being the progenitor of the eukaryotic nucleus or alternatively: The large dsDNA virus functioning as eukaryotic nucleus integrated further viral competences not being part of the large dsDNA virus. If we look at the key features of several candidates for these integration primarily we may look at prokaryotic, eukaryotic and archaeal phages:

Prokaryotic phages like cyanophages have doublestranded DNA, DNA Polymerases and RNA Polymerases similar to eukaryotes. Eubacterial phages possess linear doublestranded DNA, telomeres, DNA polymerases, RNA polymerases, chromatin and internal membranes. Archaeal phages with linear doublestranded DNA have telomere repeats similar to eukaryotes but very dissimilar to prokaryotes and they possess chromatin and a internal lipid tendency to non-lytic, persistent (and often mixed) infections (Villarreal 2005).

Other DNA-viruses share similar features which are crucial characteristics of the eukaryotic nucleus not found in prokaryotes, like Vaccinia Virus (Poxvirus) (Takemura 2001). These viruses possess a membrane-bound division of transcription and translation, multiple membranes and its DNA synthesis combines membrane loss and a cell-cycle dependent restoration as well as a actin/tubulin bound transport system (Villarreal 2005, Van Lent and Schmitt-Keichinger 2006) and nuclear pores. Cytoplasmic DNA-Viruses (African

Swine Fever Virus) have chromatin and linear chromosomes with telomeres. Phycto DNA-Virus have mRNA capping, Introns and diverse DNA replication proteins. TTV-1 to 4 have linear doublestranded DNA genomes with molecular basis for the evolution of eukaryotic chromatin and Capsids which integrates internal and external lipid proteins.

Additionally all these viruses have a self and non-self identification competence. All Viruses mark their genomes, RNAs and Proteins by different kinds of chemical modifications e.g. methylation. This marking allows the differentiation between self and non-self. Non-self may be other viruses or the host genome or host related transcripts (Villarreal 2005).

6. Evolutionary roles of viruses as natural genetic engineers

To understand the evolutionary emergence of the eukaryotic nucleus with its key features of telomeres and telomerases in the eukaryotic replication process we have to reconstruct the natural genome-editing competences of viruses (Witzany 2006) which integrated a variety of key features which we found also in eukaryotes but not in prokaryotes (although in prokaryotic viruses). As recent research in microbiology demonstrated via comparative genomics and phylogenetic analyses we have to think about life based on the crucial role of natural genome editing competences of viruses (Forterre 2001, 2002, 2005, 2006, Koonin 2006, Villarreal 2005, Tran et al. 2004).

This contradicts former concepts which focused on viruses in the light of (i) escape theories, in that they are intact or deformed genetic parasites which escaped from cellular life, (ii) or they evolved from cellular ancestors, (iii) or they are not living beings because they can't live without cellular life. From these perspectives viruses couldn't play crucial roles in the evolution of cellular life.

Interestingly phylogenetic analyses does not support DNA viruses descent from cellular life as well as they show that DNA viruses and RNA viruses most likely didn't have a common ancestor but evolved independently. It seems to be likely that we have to think about viruses at the very beginning of life, long before cellular life evolved.

6.1. Pre-cellular life

Recent research suggests to think of the early stages of life as a precellular RNA gen-pool with RNA viruses, Retroviruses, and by reverse transcriptase of singlestranded RNA viral genomes also doublestranded DNA Viruses (Leipe et al. 1999, Martin 2005, Koonin et al. 2006, Brosius 2003, Flavell 1995). Prior to cellular life forms we can imagine networks of solely chemical connected molecules coherent to the molecular syntax of RNA and latter on DNA. Several genes that are central for viral replication are missing from cellular genomes although phylogenetic analyses show that they are older than cellular elements. Overlapping arrays of unrelated viruses enshure key functions in genome replication: capsid protein, helicase superfamily in all RNA and DNA viruses (Koonin 2006).

All RNA viruses share RNA dependent RNA polymerase and reverse transcriptase which indicates a RNA virus dependent function essential for eukaryotic replication (Temin 1985), keeping in mind, that the eukaryotic nucleus seemingly derived from a large DNA virus (Bell 2001, 2006). Capsid proteins involved in jelly roll capsid protein may be a starting event in building true viruses. Alternative capsid proteins with helical capsid features might be a parallel development.

Membrane lipids, cell walls, as many other features are unrelated in bacteria and archaea. Complex colonization of unrelated viral descents into the large DNA virus which is the ancestor of the eukaryotic nucleus forced a digital/symbolic molecular grammar in the eukaryotic genome. Only by this new grammatical competence it was possible to create

diverse new features of eukaryotic cellular organization and coordination which is lacking the prokaryote world (Villarreal 2005).

6.2. Persistent viral life-strategies are beneficial for their hosts

Acute viruses that exhibit lytic action induce disease and even death. Persistent life-style of viruses implies compatible interactions with the host, either by being integrated into the hosting genome (Gorinsek et al. 2004) or within the cell plasma, and act non-destructive during most life stages of the host. The persistent life-style allows the virus to transmit complex viral phenotypes to the hosting organism. Doing so enables the host to broaden evolutive potentials for adaptational purposes and may well lead to the formation of new species. (Villarreal 2005)

6.3. Persistent status through addiction modules

The persistent status emerges through multiple colonization events into a host which neutralizes former antagonistic and incompatible features of competing viral agents without harming the host (Ryan 2004, 2006, 2007). It is important to keep in mind that this is no rare event. We have to imagine that 1 ml seawater contains 1 million bacteria but ten times more viral agents. Every bacteria is colonized by phages this means 10^{25} infections per second. This means that viruses are the most abundand life forms on earth.

Most of the genetic/genomic text editing competences which are inherent to cells, bacteria, protozoa, plants, animals, fungi are a complementary mix of former antagonistic viral features. We can identify them even today as toxin/antitoxin-, restriction/modification-, insertion/deletion – modules (Villarreal 2005, Gerdes 2000, Makarova et al. 2006). As symbiotic neutralization and counterpart regulation they represent new phenotypic features which may consist up to 100 new genes. The feature of one competence is regulated exactly by the antagonist according to developmental stages in cell-cycle, replication, tissue growth or similar contexts. Is this suppressor-function out of balance the normally downregulated part may become lytic or disease causing. Even the amplifiable and transmissible RNAi immune function indicates an old RNA-viral ancestor reaching a persistent non-lytic status which successfully excluded competing dsRNA viruses (Villarreal 2005).

The gene functions of eukaryotes acquired from persistent viruses include immunity (restriction and modification modules, toxic and antitoxic modules), silencing functions/micro-RNAs, (RNAi, methylation, suppression), recognition functions (replicate expression, receptors, expression factors) and immune regulation (signal mediating, heredity, adaptation). (Villarreal 2005)

6.4. Endogenous retroviral competences are a persistent symbiotic life-style

Endogenous retroviral competences in the persistent status are often characterized by features which are expressed only in a strictly time window of the developmental process, such as e.g., axis formation, trophectoplast formation, s-phase of the cell cycle. In this highly spezialized contexts they are replicated through signaling which blocks the suppression of the replication process. After the function is fullfilled, a signal initiates suppressor function again. The descent of retroelements with its (i) higher order regulatory functions, (ii) capability for genetic creativity and (iii) innovation competence of new regulatory patterns and combinations descended from retroviruses which can be easily identified in their three essential parts *gag*, *pol* and *env* (Rashkova et al. 2002, Weiss 2006, Tang et al. 1999). Most endogenous retroviruses have been degraded into formerly connected domains, but they can still be recognized by one of their three genes *gag*, *pol* and *env* (Gao et al. 2003, Sfakianos

and Hunter 2003, Ryan 2004, Gabus et al. 2006). The *gag* gene encodes structural proteins, *pol* encodes enzymes such as reverse transcriptase and integrase functions and *env* encodes envelope proteins. They may have important but harmloss functions within the host genome

The retroelements in general are key regulators with sophisticated competences active as natural genetic engineers (Sternberg and Shapiro 2005). Although all retroelements are related and share the gene for reverse transcriptase, there is a decisive difference between exogenous (infective) and endogenous ("defective",i.e. adapted) retroelements.

6.5. "Elements", "Entities", "Parasites" – agents of natural genome editing

Recent research shows a high abundance of dynamic DNA-remodelling by small RNAs and micro RNAs being competent in a great variety of DNA arrangements, rearrangements and recombination (Shapiro 2002, Vaughn and Martienssen 2005, Mattick 2001, 2006). Some authors speak about agents of genomic creativity (Ryan 2006), some about mobile elements (Eickbush 1999) or entities (Daubin and Ochman 2004), some about transposable elements (Slotkin and Martienssen 2007), others about mobile DNA species, or genetic parasites (Nakamura and Cech 1998, Villarreal 2005). Together these agents enable complex organisms to integrate several temporal steps and a great variety of coordinated signalling processes in eukaryotic cell replication, fix them in conserved DNA storage medium and if necessary resolve conservation, change, rearrange or newly construct parts or the whole genomic content and sequence order (Shapiro 2006).

From a biosemiotic perspective which investigates combinatorial (syntactic), content-specific (semantic) and contextual (pragmatic) rules of natural genome editing and genetic text processing it is important to notice that there is no editing without a subject that edits, i.e. an editor or a swarm of editors (Vetsigian et al. 2006) as in the spliceosome which works as integrated network of several small nuclear RNAs and its associated proteins. All these actors are acknowlegded as competent agents of DNA editing as well as catalyzing such as small RNAs sometimes connected with proteins (Vaughn and Martienssen 2005). It becomes obvious that they are authority subjects to act competent on the molecular syntax of the DNA language.

Especially without the key agents of DNA replication mRNA, tRNA and rRNA, life could not function. Not only in the case of rRNA but also in that of tRNA and the processing of the primary transcript into the pre–mRNA and the mature RNA it is obvious now that they descended from retroelements with viral origins (Eickbush and Eickbush 2007, Maizels and Weiner 1993, Maizels et al. 1999, Flavell 1995)

Although we may imagine how sophisticated the competent subject-like agents (self and non-self identification) acts in the example of endogenous retroviruses which reached persistent and non-lytic life style, and know that all related retroelements share a common genome editing competence like transposable "elements" its difficult to reconstruct the way how all these DNA encoded RNA agents reached persistent status in hundreds, thousands and tens of thousands of elements.

We only know that they act in a precise coordinated manner which would be impossible without competent signalling which includes a strict competence for self non-self identification as it is a major asset of RNAs in general and in small nucleolar RNAs in detail (Filipowicz 2000).

Persistent endogenous agents of natural genome editing seem to prefer a special kind of habitat which is characterized to be non-protein-coding DNA sectors. They use a molecular syntax which mainly consists of repeats. They colonized analog DNA genomes by inserting their sites between coding elements. This reaches so far that in the human genome only 3% of coding regions remained and 97% serves as habitat for persistent viral actors which orchestrated a highly sophisticated division of labor. From there they can regulate actively

coding sequences as they are competent to change special DNA contents throughout the whole genome. All eukaryotic DNA replication processes share a cut and paste process in which non-coding elements, introns are spliced out and remaining exons which code for proteins are put together to a coherent protein coding content ready for translation.

In difference to persistent endogenous agents of natural genome editing in eukaryotes we find persistent exogenous agents in prokaryotes competent in natural genome editing in the prokaryotic genpool. This process has long visualized as horizontal gen transfer and is now recognized as occurring by plasmids, phages and transposons all of viral descent (Frost et al. 2005).

In difference to think about mere molecules or molecule buildings being "competent" to process the sophisticated DNA language its not so difficult to think of viruses being these subject-like agents.

6.6. Eukaryotic genome content by installing a deep grammar into a superficial grammar

Higher order regulations which are performed by non-protein-coding RNAs and are inherent in all repeat elements like e.g. subtelomeric repeats and all the other retroelements have a similar relationship to proteincoding sequences as (1) deep grammar and (2) superficial grammar of utterances. Through this two different levels it is possible to determine the proteincoding data-sets according to different needs into "multiple protein meanings" (Ast 2005). In eukaryotic genome evolution the step from continous coding sequence order to interrupted sequence order occurred. Interestingly the former is characteristic for circular prokaryotic genomes whereas the latter is for linear genomes.

Telomeres themselves are not typical sites for colonization events in contrast to sites very close to these telomeres. This is similar to the phylogenetically related centromeres. Because telomeres and centromeres themselves are rather free of inverted repeats or retrolelements this could be an indicator of an ancient inherent immune (RNAi) function which protects both from invasions of genetic parasites. Biosemiotically this symbiogenetically induced invention of multiple-invaded coding data-sets by retroelements opened the possibility to use proteincoding data-sets according to various types of higher order regulation. The proteincoding data-sets are the structural vocabulary, the non-protein coding "underworld" (Mattick 2007) of RNAs are the text editing agents. In real life action this was a massive invasion of non-coding introns (viruses) into the genomic habitats of protein-coding datasets (Rogozin et al. 2005). So the molecular syntax of proteincoding data-sets could be used for different needs in different contexts (pragmatics) to serve for different genetic content arrangements (semantics):

- This could explain that in evolutionary history genetic phenotypes from one species are transferred and integrated in the genomic content of other species to get a new role or a new phenotypic feature in another context. This happened with telomeres in the linear chromosomes of ancient doublestranded DNA viruses (poxvirus, vaccinia virus, archaeal phages: AFV-1, SIRV-1, TTV 1-4) where they had other functions than in the eukaryotic genomic content (Villarreal 2005).
- This also could explain the close coherence of protein coding datasets between humans and chimpanzees (99 %) keeping in mind that the percentage of protein coding in humans and chimpanzees is only 3% whereas the percentage of non-coding DNA with higher order regulatory functions is 97% which determine coordinated expression patterns (Witzany 2006).

- This also could explain that specific cellular functions are encoded not strongly conserved at the sequence level in contrary to their preserved domains as occurs with genes of nuclear pores (Bapteste et al. 2005)
- This could open a biosemiotic perspective on the presence of a highly dynamic natural genome-editing RNA-world in preserving DNA-habitats serving as relatively stable storage medium protecting the evolutionary protocols.

7. Conclusion

Telomeric repeats protect eukaryotic genomes from degradation, repair-enzymes, genetic parasites and represent original skills of viruses. It seems likely that telomere repeats had a similar immune function in linear chromosomes of DNA viruses prior to the evolution of eukaryotes. Because repetitive elements are key features of retroviruses, I suggest an infection event of large DNA viruses by RNA viruses which attained a persistent status in the linear chromosome of the DNA virus. This protects linear chromosome ends against competing genetic parasites. The acquisition of the telomere repeats in eukaryotes has been a key event in eukaryotic nucleus evolution. The eukaryotic nucleus most likely evolved from a large doublestranded DNA virus. However, the changing structure of the eukaryotic genome with its protein-coding and non-protein-coding sections and its typical repetitive (higher-order regulatory) elements, indicate high rates of persistent, non-lytic retroviral infections. In contrast to most of these higher order regulatory agents which are mobile telomere repeats (as well as centromeres) got a conserved non-mobile status.

Telomerase from biosemiotic perspective is a natural genetic engineering tool with different functions in different contexts. Whereas reverse transcriptase has been used in RNA-Virus life cycle for replication functions, it is as acquired tool (transported and integrated in a symbiogenetic infection event getting a persistent genomic status) for complete replication of chromosomal ends in linear eukaryotic genomes. In eukaryotes telomerases and other reverse transcriptases act as endogenous retroviral competences.

In these symbiogenetic infection events the eukaryotic host acquired a higher order regulated genomic syntax which is the precondition for multiple protein meanings from the same genetic data-set through post-transcriptional modifications such as alternative splicing pathways. Therefore the transformation of the continous (prokaryotic) molecular syntax into a eukaryotic molecular syntax invaded by multiple retroelements is a major step in evolution of multicellular complexity.

From biosemiotic view all “elements”, “entities”, “agents”, “parts” being competent in natural genetic engineering and natural genome editing have phylogenetic relations to viruses.

Ackowledgement

First presented at the Cold Spring Laboratory Meeting on “Telomeres and Telomerases”, 3-6 May 2007. I want to thank Cold Spring Laboratory for participation support.

References

Ast, G. (2005). The Alternative Genome. Scientific American 292: 58-65.

Batzer, M.A. and D.L.Deininger (2002). ALU Repeats and Human Genomic Diversity. Nature Reviews Genetics 3: 370-380.

Bell, P.J.L. (2001). Viral Eukaryogenesis: Was the ancestor of the nucleus a complex DNAVirus? Journal of Molecular Evolution 53: 251-256.

Bell, P.J.L. (2006). Sex and the eukaryotioc cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. Journal of Theoretical Biology doi: 10.1016/j.jtbi.2006.05.015

Bapteste E, Charlebois RL, MacLeod D, Brochier C. (2005). The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. Genome Biology 6:R85

Bird CP, Stranger BE, Dermitzakis ET. (2006). Functional variation and evolution of non-coding DNA. Current Opinion in Genetics & Development 16:559-564.

Blasco M. (2007). The epigenetic regulation of mammalian telomeres. Nature Reviews 8: 299-309.

Boeke JD. (2003). The Unusual Phylogenetic Distribution of retrotransposons: A Hypothesis. Genome Research 13:1975-1983.

Brosius J. (2003). The Contribution of RNAs and Retroposition to Evolutionary Novelties. Genetica 118: 99-115.

Chaconas G. (2005). Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end. Molecular Microbiology 58: 625-635.

Cochrane AW, McNally MT, Mouland AJ. (2006). The retrovirus RNA trafficking granule: from birth to maturity. Retrovirology 3:18.

Coffin JM, Hughes AH, Varmus HE. (1997). Retroviruses. New York: Cold Spring Harbor Laboratory Press.

Cottingham FR, Hoyt MA. (1997). Mitotic spindle positioning in saccharomyces cerevisiae is accomplished by antagonistically acting microtubule motor proteins. Journal of Cell Biology 138:1041-1053.

Couzin J. (2002). Small RNAs make big splash. Science 298: 2296-2297.

Curcio MJ, Belfort m. (2007). The beginning of the end: Links between ancient retroelements and modern telomerases. Proceedings of the National Academy of Sciences of the USA 104: 9107-9108.

Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. EMBO Journal 21: 2746-2756.

Daubin V, Ochman H. (2004). Start-up entities in the origin of new genes. Current Opinion in Genetics & Development 14: 616-619.

Du S, Traktman P. (1996). Vaccinia virus DNA replication: Two hundred base pairs of telomeric sequence confer optimal replication efficiency on minichromosom templates. Proceedings of the National Academy of Sciences of the USA 93:9693-9698.

Eickbush TH (1997). Telomerase and retrotransposons: which came first? Science 277:911-912.

Eickbush T (1999). Mobile introns: Retrohoming by complete reverse splicing. Current Biology 9:11-14.

Eickbush TH, Eickbush DG. (2007). Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. Genetics 175: 477-485.

Eigen M, Winkler R (1975). Das Spiel – Naturgesetze steuern den Zufall. München: Piper.

Fajkus J, Sykorova E, Leitch AR. (2005). Telomeres in evolution and evolution of telomeres. Chromosome research 13: 469-479.

Filipowicz W. (2000). Imprinted expression of small nucleolar RNAs in brain: Time for RNomics. Proceedings of the National Academy of Sciences of the USA 97: 14035-7.

Fire A, Xu SQ, Montgomery MK, Kostas SA, driver SE, Mello CC. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391: 806-811.

Flavell AJ. (1995). Retroelements, reverse transcriptase and evolution. Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology 110: 3-15.

Forterre P. (2001). Genomics and early cellular evolution. The origin of the DNA world. Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie 324: 1067-1076.

Forterre P. (2002). The origin of DNA genomes and DNA replication proteins. Current Opinion in Microbiology 5: 525-532.

Forterre P. (2005). The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 87: 793-803.

Forterre P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. Virus Research 117: 5-16.

Forterre P. (2006). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: A hypothesis for the origin of cellular domain. Proceedings of the National Academy of Sciences of the USA 103: 3669-3674.

Frost LS, Laplae R, Summers AO, Toussaint A. (2005). Mobile Genetic Elements: The Agents of Open Source Evolution. Nature Reviews Microbiology 3: 722-732.

Gabus C, Ivanyi-Nagy R, Depollier J, Bucheton A, Pelisson A, Darlix JL. (2006). Characterization of a nucleocapsid-like region and of two distinct primer tRNA binding sites in the endogenous retrovirus Gypsy. Nucleic Acids Research 34: 5764-5777.

Gao X, Havecker ER, Baranov PV, Atkins JF, Voytas DF. (2003). Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA 9: 1422-1430.

Gorinsek B, Gubensek F, Kordis D. (2004). Evolutionary Genomics of Chromovirus in Eukaryotes. Molecular Biology and Evolution 21:781-798.

Greber UF. (2002). Signalling in viral entry. Cell Molecular Life Sciences 59:608-626.

Grewal SIS, Elgin SCR. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature 447: 399-406.

Haoudi A, Mason JM. (2000). Reverse transcriptase can stabilize or destabilize the genome. Genome 43: 949-956.

Ijdo JW, Baldini A, Ward DC, Reeders ST, Wells RA. (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. Proceedings of the National Academy of Sciences of the USA 88: 9051-9055.

Jady BE, Bertrand E, Kiss T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. The Journal of Cell Biology 164: 647-652.

Kiss AM, Jady BE, Darzaq X, Verheggen C, Bertrand E, Kiss T. (2001). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids research 30: 4643-4649.

Koonin EV, Senkevich TG, Dolja VV. (2006). The ancient Virus World and evolution of cells. Biology Direct 1:29.

Koonin EV (2006). Temporal order of evolution of DNA replication system inferred by comparison of cellular and viral DNA polymerases. Biology Direct 1:39, doi:10.1186/1745-6150-1-39

Laun P, Bruschi CV, Dickinson JR, Rinnerthaler M, Heeren G, Schwimbersky R, Rid R, Breitenbach M. (2007). Yeast mother cell-specific ageing, genetic (in)stability, and the somatic mutation theory of ageing. Nucleic Acids Research doi:10.1093/nar/gkm919, 1-14.

Leipe DD, Aravind L, Koonin EV. (1999). Did DNA replication evolve twice independently? Nucleic Acids research 27: 3389-3401.

Maita N, Anzai T, Aoyagi H, Mizuno H, Fujiwara H. (2004). Crystal Structure of the Endonuclease Domain Encoded by the Telomere-specific Long Interspersed Nuclear Element, TRAS1. Journal of Biological Chemistry 279: 41067-41076.

Maizels A, Weiner AM. (1993). The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication. In: The RNA World. Gesteland RF, Atkins JF (Eds). Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, pp 577-602.

Maizels N, Weiner AM, Yue D, Shi P. (1999). New Evidence for the genomic Tag Hypothesis: Archaeal CCA-Adding Enzymes and tRNA Substrates. Biological Bulletin 196: 331-334.

Makarova KS, Grishin NV, Koonin EV. (2006). The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. Bioinformatics 22:2581-2584.

Martin W. (2005). Archaebacteria (Archaea) and the origin of the eukaryotic nucleus. Current Opinion in Microbiology 8: 630-637.

Matera AG. (2006). Drosophila Cajal bodies: accessories not included. The Journal of Cell Biology 172: 791-793.

Matera AG, Terns RM, Terns MP. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nature Reviews Molecular Cell Biology 8: 209-220.

Mattick JS. (2001). Non-coding RNAs: the architects of eukaryotic complexity. EMBO Reports 2: 986-991.

Mattick JS. (2006). The underworld of RNA. Nature Genetics 38: 393.

Mesnard JM, Lebeurier G. (1991). How do viral reverse transcriptases recognize their RNA genome? FEBS Letters 287:1-4.

Nakamura TM, Cech TR. (1998). Reversing Time: Origin of Telomerase. Cell 92: 587-590.

Nosek J, Kosa P, Tomaska L. (2006). On the origin of telomeres: a glimpse at the pre-telomerase world. Bioessays 28: 182-190.

Platani M, Goldberg I, Lamond AI, Swedlow JR. (2002). Cajal Body dynamics and association with chromatin are ATP dependent. Nature Cell Biology 4: 502-508.

Rashkova S, Karam SE, Kellum R, Pardue ML. (2002). Gag proteins of the two drosophila telomeric retrotransposons are targeted to chromosome ends. The Journal of Cell Biology 159: 397-402.

Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV. (2005). Analysis of evolution of exon-intron structure of eukaryotic genes. Briefings in Bioinformatics 6: 118-134.

Ryan, F.P. (2004). Human endogenous retroviruses in health and disease: a symbiotic perspective. Journal of the Royal Society of Medicine 97: 560-565.

Ryan FP. (2006). Genomic creativity and natural selection: a modern synthesis. Biological Journal of the Linnean Society, 2006,88: 655–672

Ryan FP. (2007). Viruses as symbionts. Symbiosis 44: 11-21.

Sfakianos JN, Hunter E. (2003). M-PMV capsid transport is mediated by Env/Gag interactions at the pericentriolar recycling endosome. Traffic 4:671-680.

Shabalina SA, Spiridonov NA. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biology 5: 105e.

Shapiro JA. (2002). Genome Organization and Reorganization in Evolution. Annals of the New York Academy of Sciences 981: 111-134.

Shapiro JA. (2006). Genome Informatics: The Role of DNA in Cellular Computations.

Biological Theory 1:288-301.
Shapiro JA, Sternberg R. (2005). Why repetitive DNA is essential to genome function. Biological Reviews 80: 1-24.
Slotkin RK, Martienssen R. (2007). Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8: 272-285.
Sternberg R. (2002). On the Roles of Repetitive DNA Elements in the Context of a Unified Genomic-Epigenetic System. Annals of the New York Academy of Sciences 981: 154- 188.
Sternberg, R. and J.A. Shapiro (2005). How repeated retroelements format genome function. Cytogenetic and Genome Research 110: 108-116.
Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SIS. (2005). RNA-dependent RNA polymerase is an essential component of a self-inforcing loop coupling heterochromatin assembly to siRNA production. Proceedings of the National Academy of Sciences of the USA 102: 151-157.
Takemura M. (2001). Poxviruses and the origin of the eukaryotic nucleus. Journal of Molecular Evolution 52: 419-425.
Tang Y, Winkler U, Fredd EO, Torrey TA, Kim W, Li H, Goff SP, Morse HC. (1999). Cellular Motor Protein KIF-4 associates with retroviral Gag. Journal of Virology 73:10508-10513
Temin HM. (1985). Reverse Transcription in the Eukaryotic Genome: Retroviruses, Pararetroviruses, Retrotransposons and Retrotranscripts. Molecular Biology and Evolution 2: 455-468
Tomlinson RL, Ziegler TD, Supakorndej T, Terns RM, Terns MP. (2006). Cell Cycle-regulated Trafficking of Human Telomerase to Telomeres. Molecular Biology of the Cell 17: 955-965.
Tourand Y, Bankhead T, Wilson SL, Putteet-Driver AD, Barbour AG, Byram R, Rosa PA, Chaconas G. (2006). Differential Telomere Processing by Borrelia Telomere Resolvases In Vitro but Not In Vivo. Journal of Bacteriology 188: 7378-7386.
Tran E, Brown J, Maxwell ES. (2004). Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus? Trends in Biochemical Sciences 29: 343-350.
Vale R. (2003). The Molecular Motor Toolbox for Intracellular Transport. Cell 112: 467-480.
Van Lent JWM, Schmitt-Keichinger C. (2006). Viral Movement Proteins Induce Tubule Formation in Plant and Insect Cells. In: F. Baluska, d. Volmann, P. Barlow (eds) Cell-Cell Channels. Eurekah.com, 1-13
Vaughn MW, Martienssen R. (2005). It's a Small RNA World, After All. Science 309: 1525-15-26.
Villarreal LP. (2005). Viruses and the Evolution of Life. Washington, ASM Press.
Villasante A, Abad JP, Mendez-Lago M. (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proceedings of the National Academy of Sciences of the USA 104: 10542-10547.
Volff JN. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28: 913-922.
Weber MJ. (2006). Mammalian Small Nucleolar RNAs are mobile genetic elements.
Weiss RA. (2006). The discovery of endogenous retroviruses. Retrovirology 3: 67. doi: 10.1186/1742-4690-3-67.
Witzany G. (2006). Natural Genome-Editing Competences of Viruses. Acta Biotheoretica 54: 235-253.
Xiong Y, Eickbush TH. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. The EMBO Journal 9: 3353-3362.

Yang J, Malik HS, Eickbush TH. (1999). Identicication of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. Proceedings of the National Academy of Sciences of the USA 96: 7847-7852.

Zemann A, Beckke A, Kiefmann M, Brosius J, Schmitz J. (2006). Evolution of small nucleolar RNAs in nematodes. Nucleic Acids Research 34: 2676-2685.